CRAVATH, SWAINE & MOORE LLP

WORLDWIDE PLAZA
825 EIGHTH AVENUE
NEW YORK, NY 10019-7475

TELEPHONE: +1-212-474-1000
FACSIMILE: +1-212-474-3700

CITYPOINT
ONE ROPEMAKER STREET
LONDON EC2Y 9HR
TELEPHONE: +44-20-7453-1000
FACSIMILE: +44-20-7860-1150

WRITER'S DIRECT DIAL NUMBER
+1-212-474-1080

WRITER'S EMAIL ADDRESS
bgruenstein@cravath.com

JOHN W. WHITE
EVAN R. CHESLER
RICHARD W. CLARY
STEPHEN L. GORDON
ROBERT H. BARON
DAVID MERCADO
CHRISTINE A. VARNEY
PETER T. BARBUR
THOMAS G. RAFFERTY
MICHAEL S. GOLDMAN
RICHARD HALL
JULIE A. NORTH
ANDREW W. NEEDHAM
STEPHEN L. BURNS
KATHERINE B. FORREST
KEITH R. HUMMEL
DAVID J. KAPPOS
DANIEL SLIFKIN
ROBERT I. TOWNSEND, III
WILLIAM J. WHELAN, III
PHILIP J. BOECKMAN
WILLIAM V. FOGG
FAIZA J. SAEED
RICHARD J. STARK

THOMAS E. DUNN
MARK I. GREENE
DAVID R. MARRIOTT
MICHAEL A. PASKIN
ANDREW J. PITTS
MICHAEL T. REYNOLDS
ANTONY L. RYAN
GEORGE E. ZOBITZ
GEORGE A. STEPHANAKIS
DARIN P. MCATEE
GARY A. BORNSTEIN
TIMOTHY G. CAMERON
KARIN A. DEMASI
DAVID S. FINKELSTEIN
DAVID GREENWALD
RACHEL G. SKAISTIS
PAUL H. ZUMBRO
ERIC W. HILFERS
GEORGE F. SCHOEN
ERIK R. TAVZEL
CRAIG F. ARCELLA
DAMIEN R. ZOUBEK
LAUREN ANGELILLI
TATIANA LAPUSHCHIK

ALYSSA K. CAPLES
JENNIFER S. CONWAY
MINH VAN NGO
KEVIN J. ORSINI
MATTHEW MORREALE
JOHN D. BURETTA
J. WESLEY EARNHARDT
YONATAN EVEN
BENJAMIN GRUENSTEIN
JOSEPH D. ZAVAGLIA
STEPHEN M. KESSING
LAUREN A. MOSKOWITZ
DAVID J. PERKINS
JOHNNY G. SKUMPIJA
J. LEONARD TETI, II
D. SCOTT BENNETT
TING S. CHEN
CHRISTOPHER K. FARGO
KENNETH C. HALCOM
DAVID M. STUART
AARON M. GRUBER
O. KEITH HALLAM, III
OMID H. NASAB
DAMARIS HERNÁNDEZ

JONATHAN J. KATZ
MARGARET SEGALL D'AMICO
RORY A. LERARIS
KARA L. MUNGOVAN
NICHOLAS A. DORSEY
ANDREW C. ELKEN
JENNY HOCHENBERG
VANESSA A. LAVELY
G.J. LIGELIS JR.
MICHAEL E. MARIANI
LAUREN R. KENNEDY
SASHA ROSENTHAL-LARREA
ALLISON M. WEIN

SPECIAL COUNSEL
SAMUEL C. BUTLER

OF COUNSEL
MICHAEL L. SCHLER

January 29, 2019

Floyd, et al. v. City of New York, 08-CV-1034 (AT)
Ligon, et al. v. City of New York, et al., 12-CV-2274 (AT)
Davis, et al. v. City of New York, et al., 10-CV-0699 (AT)

Dear Judge Torres:

I am counsel to the Monitor, Peter L. Zimroth, and respectfully submit this letter in connection with the pending dispute related to the pilot-program proposal in the above-captioned cases. On July 19, 2018, this Court ordered the creation of a proposal for a pilot program to study the electronic recording of first- and second-level police-citizen encounters. (ECF No. 619.[1]) On August 9, 2018, this Court further ordered that the proposed program study the use of body-worn cameras ("BWCs") in first-level encounters. (ECF No. 634.) The Monitor submitted the ordered proposal (the "Proposal") on November 9, 2018. (ECF No. 660-1.) The parties and Communities United for Police Reform ("CPR") have since submitted letters related to the Proposal. The Monitor respectfully requests that this Court adopt the version of the Proposal

[1] All ECF cites refer to documents as numbered in the *Floyd* docket.

attached as Exhibit 1 to this letter, which reports results in odds ratios and clarifies that officers will be required to document encounters in a timely manner, as discussed below.[2]

1. A Brief Account of the Proposal.

The Proposal is designed to study how electronic documentation and the use of BWCs affect the legality of police-citizen encounters. (ECF No. 660-1 at 1.) It accomplishes this by organizing a sample of precincts and police service areas ("PSAs") into four groups. (*Id.* at 2-3.) The first group will follow existing police procedures, meaning that it will not document Level 1 and 2 encounters or activate BWCs during Level 1 encounters; this is the control group. (*Id.*) The second group will activate BWCs during Level 1 encounters (in addition to situations in which BWC activation is currently required, including Level 2 encounters), but will not document Level 1 or 2 encounters. (*Id.*) The third group will activate BWCs only when doing so is currently required (not during Level 1 encounters), but will document Level 1 and 2 encounters. (*Id.*) And the fourth group will both activate BWCs during Level 1 encounters and document Level 1 and 2 encounters. (*Id.*)

Officers in the third and fourth groups will document all of their Level 1 and 2 encounters for a given shift on an electronic form on their phones or tablets. (*Id.* at 5.) This form will demand more information for those encounters than the Facilitator suggested. The Facilitator recommended that officers "click and enter the approximated age, gender, race and ethnicity of any person they approach at either Level 1 or Level 2,

[2] Exhibit 2 tracks the changes between the updated Proposal and the Proposal as it was submitted on November 9, 2018. Exhibit 1 is a clean version of the updated Proposal.

and then click if the encounter escalates to Level 3," and suggested that this "could be accomplished in a matter of seconds." (ECF No. 597 at 231-32.) The Monitor's experts determined that this amount of information would not be sufficient to obtain meaningful results. For this reason, the study will ask officers to record additional information: the reason for the encounter, the event type, the ICAD number, whether the encounter was recorded by a BWC or not, the call type, and the officer's initial and final levels of suspicion for each encounter. (ECF No. 660-1 at 5.) The Proposal takes into account that this level of recording will be more time-consuming and burdensome—it will likely take more than "a matter of seconds." As noted below in Section 4, the Proposal requires officers to document their encounters in a timely manner.

In all four groups, trained observers will ride along with a sample of randomly selected officers and record the actions of the officers and the circumstances of each encounter using a protocol that will be developed by the Monitor team in consultation with the parties. (*Id.* at 2-3.) Importantly, unobserved officers in each studied precinct or PSA must comply with the study requirements imposed on that precinct or PSA. (*Id.*) So if a precinct is in the second group, for example, every officer within that precinct must activate her BWC during Level 1 encounters regardless of whether she is accompanied by an observer or not. (*Id.*) Therefore, the same officers will be following the study requirements both with and without observation. The Proposal examines the effects of observer presence by comparing the behavior of officers while under observation with their behavior while unobserved. (*Id.* at 9, 14.)

The Proposal expects that the observers will witness approximately 1,512 encounters. (*Id.* at 4.) As noted on page 4 of the Proposal, it would be reasonable to

estimate that there will be an average of 12 police-citizen contacts per eight-hour shift, based on prior systematic-observation studies and consultation with the NYPD. (*Id.*) However, to be more conservative, the Proposal assumes a rate of nine citizen contacts per shift. With 168 observation sessions, that would result in a sample size of 1,512 observed contacts. If the rate of citizen contacts is closer to 12 per shift, the sample size will be even larger.[3]

Each encounter will be attributed to the officer team involved in the encounter as a team (e.g., an officer and his partner, or a team of officers in an anti-crime unit), not to individual officers. (*Id.* at 3-4.) Thus, each encounter will be attributed to a two-officer uniformed patrol, a two-officer PSA unit, or a multi-officer anti-crime unit. (*Id.*)

The reports of the observers will go to the Monitor team, which will assess the legality of the encounters described in them. (*Id.* at 6.) The Monitor team will make these assessments using a protocol that will take its criteria from the standards in the court-approved NYPD policies, on which NYPD officers are educated in the court-approved Stop, Question, and Frisk ("SQF") training.

[3] Plaintiffs' experts initially suggested that the study should have a sample size of about 1,290 encounters to achieve sufficient statistical power for the analysis proposed. The Monitor team suggested going further and using a sample of 1,512 encounters—and Plaintiffs' experts agreed to that sample size. It thus came as a surprise to the Monitor team that Plaintiffs' experts now believe the size of the study is insufficient. Moreover, Plaintiffs' experts do not propose an alternative size and, in any case, as described in the next section, the duration of the study can be extended if it is determined that the sample size is too low.

Once the Monitor team's assessments are complete, the team's experts will use those assessments to determine whether electronic documentation and BWCs had statistically significant positive impacts on the legality of encounters.

2. The Proposal Does Not Need to Be Changed at This Time to Enhance Its Statistical Capabilities.

Plaintiffs argue that the Proposal should be changed to increase its statistical power before the study begins. (ECF No. 669 at 1-2.) As explained more fully below, statistical power is the probability that the study will demonstrate a difference in conditions that is statistically significant. To the extent that each precinct's officers behave in specific ways that differ from the behaviors of officers of other precincts, the observed encounters in each precinct will not be "independent" of one another. This will decrease the Proposal's "effective" sample size and, thus, its statistical power, making it more difficult to say with confidence that differences among the conditions are due to the policies being tested rather than chance.

At this time, however, we do not know the extent to which officers in each precinct will behave distinctively. We do know that, using reasonable assumptions about the amount of non-independence we are likely to observe, the Proposal is sufficiently powered. The Monitor team's experts will be monitoring the progress of the study and, if it appears that non-independence will be higher than forecasted and the study's effective sample size will be too low, the duration of the study will be extended appropriately to incorporate more observed encounters and raise the sample size. Extending the duration of the study once it is underway has no different effect on the study than committing to extend the study now. But extending the study now could prove wasteful, expending

time and resources better devoted to designing and implementing any reforms recommended by the study.

A. The Basics of Statistical Power and Independence.

Statistical power represents a study's ability to demonstrate a difference in conditions that is statistically significant. To illustrate this, suppose we have altered a coin to land on heads more often than tails when flipped, and we want to know whether our design succeeded. When we flip the loaded coin 100 times, it lands on heads 90 times; when we flip a normal coin 100 times, it lands on heads 50 times. Statistical significance represents the likelihood that the difference between 90 and 50 is attributable to the design of the loaded coin rather than chance. In other words, it is possible—but unlikely—for a normal coin to come up heads 90 times if we flip it 100 times. Using the standard of most rigorous scientific research, we would say there is a statistically significant difference between the loaded and normal coins if the chance that our "loaded coin" is really just a normal coin that happened by mere coincidence to land on heads 90 times is 5% or less.

The probability that a study has the power to demonstrate that statistically significant difference between the coins, in turn, is statistical power. Suppose that we give the coins to someone else and ask her whether one is loaded. She flips each coin 10 times, and the loaded one lands on heads nine times whereas the normal one lands on heads five times. This study seems to show the same relative difference between the coins as the 100-flip study. But this study has much less power to say that the chance is less than 5% that the difference between the coins is attributable to one being loaded. In other words, the odds that a normal coin will land on heads nine times out of 10 from

mere chance are much greater than the odds it will do so 90 times out of 100. The standard of most rigorous scientific research is that a study should have at least an 80% chance of demonstrating a difference like that between the coins at a statistically significant level. Because the 10-flip study cannot demonstrate a statistically significant difference between the coins—the study cannot demonstrate that there is a chance less than 5% that the loaded coin is different than the normal coin—the study is not adequately powered.

In general, then, a study's statistical power increases as its sample size increases. But another consideration in a study's statistical power is the "independence" of the observations within its sample. To illustrate this, imagine we want to test the hypothesis that as weather gets warmer, attendance at Major League Baseball ("MLB") games increases. A study of eight games could not yield a statistically significant conclusion, and so would not be sufficiently powered. Suppose then we took a sample of 80 games, with 10 games per team and eight teams. Then we have a question of statistical independence: should our study count these as 80 games, which could generate adequate power, or as eight? The answer depends on how independent each team's games are from one another. Suppose that each team's fan base displays a distinctive pattern: each time the temperature increases by five degrees, the first team's attendance increases by 1,000 people, the second team's attendance increases by 1,100 people, and so on. In that case, we do not want to draw a conclusion about the pattern in the MLB as a whole because all we have as evidence is the distinctive patterns of eight teams, rather than 80 games that are representative of the whole league. In other words, because each team's games are not "independent" of one another, we count all of each team's games

once: we have what is called an "effective" sample size of 8, not 80. In light of this reduction in sample size, the study is insufficiently powered.

But suppose instead that there is no distinctive pattern for each team and rather a pattern for all teams: every time the temperature increases by five degrees, the attendance for *every* team increases by 1,000 people. In that case, we have no reason to believe that our choice of teams affects the ability of the sample to represent the MLB as a whole. Thus, it makes sense to count the 80 games in our sample as independent—80 separate observations (an "effective" sample size of 80), which might make for an adequately powered study. Accordingly, it is important to know not just the "actual" size of the sample, but also how the non-independence between the observations in the sample impacts the "effective" sample size.

B. The Proposal Does Not Need to Increase Its Projected Effective Sample Size at This Time.

Plaintiffs and Professors Jeffrey Fagan and Jack Glaser argue that the Proposal is insufficiently powered on the ground that there is too much non-independence among the Proposal's anticipated observations. (ECF No. 669 at 1-2; ECF No. 669-1 at 1-2.) First, they observe that each studied police-citizen encounter will not necessarily reflect the judgment of one observed officer. That officer may create an encounter at the behest of another officer, for example. This could generate non-independence if that encounter is attributed to both officers even though only one made the decision about whether to initiate or how to code the encounter. Alternatively, it would be inaccurate to attribute the encounter just to the observed officer if the other made the decisions. (*See* ECF No. 669-1 at 1.) Second, Professors Fagan and Glaser observe that the Proposal

counts multiple encounters by the same observed officer unit multiple times. Five encounters by Officer Unit X count as five observations, even though those encounters all reflect Officer Unit X's particular training and circumstances and so are not independent: those encounters are "clustered" or "nested" within Officer Unit X. (*Id.* at 1-2.) Third, Professors Fagan and Glaser observe that the Proposal counts separate encounters by officers within the same precinct separate times—even though those officers may have similar training and circumstances, which makes the encounters somewhat non-independent. (*Id.* at 2.) None of these critiques warrants any change to the Proposal design.

The first critique misunderstands the way that the Proposal counts encounters. That critique assumes that either an encounter by a two-officer unit will be counted twice (once for each officer) or that the encounter will be attributed to just one of the officers. Instead, the Proposal counts each encounter by a two-officer unit just once, and attributes that encounter to the unit as a whole. All encounters will be attributed to either a two-officer uniform patrol, a two-officer PSA unit, or a multi-officer anti-crime unit. (*See* ECF No. 660-1 at 4.) Thus, there is no risk that a single decision will be counted multiple times, nor that a decision will be attributed to the wrong person.

The second and third critiques do not demonstrate that the effective sample size of the Proposal is too small at this time. Rather, the Proposal's sample size, and its projected "effective" sample size, are more than adequate on the basis of reasonable estimates of the non-independence that the study is likely to encounter in the data.

The Proposal expects to observe 168 shifts, each with an average of nine encounters, so that the projected total number of observed encounters is roughly 1,512. (ECF No. 660-1 at 4.) An equal number of shifts (42) will be observed to test the effect of each of the four studied conditions, meaning that there will be roughly 378 encounters per condition. (*Id.*) Furthermore, the observed officers will be part of different administrative units—three precincts spread across all four conditions, and one PSA patrol for each condition. (*Id.*) It is also reasonable to assume that plainclothes anti-crime units within a precinct behave differently (and thus independently) from the patrol officers in that same precinct, and thus count their precincts as separate administrative units. Thus, there will be seven administrative units per condition (the uniformed-patrol part of each of the three precincts, the anti-crime-team part of each of the three precincts, and one PSA patrol).

In Plaintiffs' second and third critiques, they assert that two kinds of non-independence will erode the power of the Proposal: (1) non-independence within shifts (i.e., the nine observed encounters by the same two-officer unit in each shift are not perfectly independent) and (2) non-independence within administrative units (i.e., the observed encounters of all the units within each precinct or PSA are not perfectly independent). The amount of non-independence in these data sets is measured using a number known as the intra-cluster correlation ("ICC"). An ICC of 1 would mean that all of the encounters within shifts or administrative units conform to specific patterns that are exactly the same (perfect non-independence), whereas an ICC of 0 would mean that a shift and administrative unit are not distinctive and the encounters within them should be treated as separate observations (perfect independence). Using the ICC, we can calculate

the extent to which the "actual" sample size should be reduced after taking into account the non-independence by shift and administrative unit, creating an "effective" sample size. (Just as in the MLB example above, where the "actual" sample size of 80 games was reduced to an "effective" sample size of 8 when the games of each team were not independent.) The standard formula for performing this calculation is the following:[4]

$$N_{Effective} = (n) / (1 + (m - 1) * r + (k - m) * r)$$

"M" in this formula is the number of observed stops per officer for a shift, "k" is the number of observations per administrative unit, and "r" is the ICC.

We cannot know the extent to which shifts and administrative units will prove to be distinctive until the study begins. And the ICCs of police behavior have not been calculated or published in prior studies. But we do know that educational researchers have estimated ICCs between .10 and .25 for the behavior of students in the same class within the same school.[5] This range is consistent with the ICCs that the Monitor team's experts have seen in their research, where an ICC of .25 is considered very high. We assume that this ICC range supplies a reasonable estimate for the behavior of each two-officer unit in encounters on the same shift—i.e., we assume that there is

[4] Sally M. Kerry & J. Martin Bland, *Sample Size in Cluster Randomization*, 316 BMJ 549 (1998).

[5] *See* Larry V. Hedges & Eric C. Hedberg, *Intraclass Correlation Values for Planning Group-Randomized Trials in Education*, 29 Educ. Evaluation and Policy Analysis 60 (2007); David M. Murray et al., *Intraclass Correlation Among Common Measures of Adolescent Smoking: Estimates, Correlates, and Applications in Smoking Prevention Studies*, 140 Am. J. of Epidemiology 1038 (1994); Nichola Shackleton et al., *Intraclass Correlation Values for Adolescent Health Outcomes in Secondary Schools in 21 European Countries*, 2 SSM Population Health 217 (2016).

some, but not perfect, distinctiveness in those encounters. We further assume that an ICC range of .01 to .10 is a reasonable estimate for the administrative units, expecting that there will be somewhat less distinctiveness in encounters across each precinct than in the encounters of each officer-unit in a particular shift.

Those assumptions yield Table 1 below, which shows the "effective" sample size (based on the anticipated actual sample size of 1,512) for the range of assumed ICCs. For each of those "effective" sample sizes, Table 1 provides the minimum detectable effect size ("MDES"). The MDES represents the smallest statistically significant effect that a study can detect with adequate statistical power. So, in the example of the loaded coin, flipping a coin 100 times and coming up with heads 90 times yields a statistically significant conclusion that the coin is loaded, because 90 is so much greater than the number of heads we expect from a normal coin (50). But if the number of heads is only 60, it is very hard to say whether the coin is slightly loaded, or just a normal coin that happened by chance to land on heads 60 times. Yet the difficulty of saying with statistical significance that the coin is loaded—the inability of the study to detect that small effect size with adequate statistical power—diminishes as the sample size increases. If we flip the coin 1,000 times and it lands on heads 600 times, we can confidently conclude that the coin is slightly loaded in a way that we could not by flipping the coin just 100 times.

Accordingly, Table 1 below shows that when the "effective" sample size is largest, the Proposal has the most ability to detect small differences (MDES) between

the four study conditions at 80% statistical power.[6] But even the largest MDES in Table 1 shows the ability to detect relatively small differences. An MDES of .127 is considered a very small effect size, and an MDES of .243 is considered a medium effect size.[7] Thus, Table 1 shows that the Proposal should be able to detect the effects of the proposed reforms, most importantly in reducing unconstitutional encounters or enhancing the NYPD's ability to identify encounters that escalate to Level 2 or 3, with scientific rigor: 80% statistical power.

[6] The MDES figures in Table 1 were calculated using a general linear model formula for testing differences across four groups, in the statistical software package known as STATA (version 15.0). Those figures represent the smallest statistically significant effect size that can be detected at each given effective sample size at 80% power. *Compare* Ralph G. O'Brien & Keith E. Muller, *Unified Power Analysis for T-Tests Through Multivariate Hypotheses*, *in* Applied Analysis of Variance in the Behavioral Sciences 297 (Lynne K. Edwards ed., 1993).

[7] Jacob Cohen, Statistical Power Analysis for the Behavioral Sciences (rev. ed. 1977).

Table 1. Minimum Detectable Effect Sizes for Cluster Design of Four Conditions at Different Levels of Intra-Cluster Correlation.

ICC Shift	ICC Admin Unit	Effective sample size overall	MDES	Odds Ratio	Power to detect difference across four conditions[8]
.10	.01	672	.127	1.26	.80
.15	.025	455	.156	1.32	.80
.20	.05	312	.189	1.40	.80
.25	.075	237	.216	1.48	.80
.30	.10	191	.243	1.55	.80

In sum, the Proposal will be able to detect adequately small effects with adequate statistical power based on a reasonable forecast of the non-independence in the observed encounters. Plaintiffs did not provide any calculations to the contrary. If the non-independence turns out to be greater than forecasted, the duration of the study can be extended. But extending the study at this time is unnecessary and likely to cause undue delay in implementing any reforms recommended by the study.

3. The Proposal Will Express Effect-Size Results as Odds Ratios and Standard Deviations.

Professors Fagan and Glaser criticize the Proposal because it uses standard deviations to express the effects detected by the study's statistical analysis. They ask that

[8] This column represents the study's power to detect differences that exist between all four study groups—for example, the study's ability to detect that heightened BWC and documentation requirements have a significant impact on encounter legality. We also calculated the study's power to detect differences across two conditions—for example, the study's ability to detect that a heightened BWC requirement improves encounter legality significantly more than a heightened documentation requirement does (comparing group 2 to group 3). This was done by putting each pairing of MDES and effective sample size from Table 1 into a STATA formula for testing differences between two groups (similar to the formula for four groups discussed in note 6). For each pairing there is more than 80% power to detect the given MDES across two conditions, ranging from 89% to 84% power.

the effects be expressed as odds ratios, which they believe are easier to understand than standard deviations. (ECF 669-1 at 2.) These are two ways of expressing the same thing; using odds ratios does not change the Proposal at all. There are standard formulas, well-known to experts, for converting standard deviations into odds ratios.[9] Therefore, the Monitor team agrees that the study will use both forms of expression, as Table 1 does in the columns for MDES and odds ratios above. Both forms of expression will be used in the study results.

4. The Proposal's Clarified Requirements Regarding Documentation Are Appropriate.

Plaintiffs argue for three changes to the Proposal's requirements regarding the documentation of Level 1 and 2 encounters: (1) each encounter should be documented on a separate electronic form, rather than an electronic log with other encounters; (2) officers should record each encounter "during or immediately after the encounter itself, absent exigent circumstances"; and (3) officers should use their phones' time-and-location functionality to record the time and location of each encounter. (ECF No. 669 at 3.) They contend that these changes are required by this Court's order for the Proposal. (*Id.* at 2.) They further claim that allowing officers to wait to complete the records of their encounters would create an unjustified risk that those officers will misremember the encounters. (*Id.*) And it would also create an unjustified risk that those officers will misreport the burden of recordkeeping, Plaintiffs claim, because they are not

[9] *E.g.*, The Campbell Collaboration, Campbell Methods Series: Converting Between Effect Sizes (Dec. 13, 2016), https://campbellcollaboration.org/media/k2/attachments/converting_between_effect_sizes.pdf (explaining that one standard conversion formula is Odds Ratio = $e^{\text{effect size}*(\pi/\sqrt{3})}$).

necessarily recording under the eyes of study observers. (*Id.*) I address each of these arguments below.

This Court ordered the creation of a pilot program "to study the electronic recording of first- and second-level police-citizen encounters". (ECF No. 619 at 2.) The object of the study is to understand "whether the benefits of recording lower-level encounters outweigh the financial, administrative, and other costs". (*Id.* at 3.) The Court did not say that the study must contain all of the Facilitator's suggestions verbatim, and instead provided a set of general guidelines for what the Proposal must include in planning to reach the ultimate cost-benefit conclusion. (*Id.* at 2.)

Plaintiffs' first proposed change, a requirement that officers document each encounter on a separate electronic form, is premature. The NYPD has not finished developing the application that officers will use to record information about encounters on their phones. The application's design may ultimately require officers to document each encounter on a separate form, which would make Plaintiffs' suggestion moot.[10] But the application's development should not be impaired by an arbitrary requirement that encounters be documented on separate forms; field testing and the experts' views about how best to implement the application could, for example, prompt a consolidated log for all encounters. Plaintiffs have not explained any benefit that would accrue from requiring a separate form for each encounter, and it is not apparent why there would be any.

[10] The Proposal's use of the term "Investigative Encounters Worksheet" on pages 2 and 3 may have created the misimpression that the application has been completed and will use a particular kind of log. The updated Proposal attached as Exhibits 1 and 2 eliminates this ambiguity.

Plaintiffs' second proposed change, a requirement that officers document each encounter immediately absent exigent circumstances, is adequately captured by the clarifying additions to the Proposal as attached in Exhibits 1 and 2. Under the Proposal, officers will be required to document each Level 1 and Level 2 encounter at a scene within 30 minutes of leaving that scene. (*See* Ex. 1 at 5.) The Proposal thus provides for officers documenting Level 1 and Level 2 encounters in a timely fashion, consistent with how officers operate in the field: the time immediately after officers leave a scene provides a natural and expeditious point at which to document.

Officers will be instructed that only a limited set of substantive reasons can justify them in not documenting within 30 minutes of leaving the scene. (*See id.*) These reasons will appear in a drop-down menu in the electronic documentation system; to the extent that one of those reasons reasonably requires the officer to delay documentation of an encounter, she will be obligated to select that reason in the menu when she is able to document. (*Id.*) In this way, the Proposal ensures that documentation will generally be done at or soon after each encounter, and that officers articulate and record substantive reasons for doing otherwise.

Plaintiffs' proposed third change is not justified. Using phones' time-and-location functionality to record the time and location of each encounter will not provide the most accurate data. That is because officers will often need to document encounters in a different time or place than they occurred. For example, officers frequently arrive at scenes where multiple civilians must be interviewed and have no time to document the first interview until the last one is completed. And an officer often begins travelling to her next job immediately after the last has occurred. Requiring the officer to complete

the record of an encounter that concluded minutes earlier while stationary rather than en route (as would a requirement that she record location using her phone's functionality) would promote inefficiency without enhancing her memory of that encounter. Most importantly, when officers must delay documentation for substantive reasons, their phones' functionality will obviously not provide accurate time and location data. The better solution for accurate data is the existing one, requiring officers to enter the time and location of their encounters as part of documenting them. These records can be checked against those of the independent observers in the study, riding along with the officers, who will be recording the times and locations of the encounters they witness. (*See* ECF No. 660-1 at 7.)

Finally, observers' roles should not, as Plaintiffs suggest, include documenting how burdensome it is for officers to complete their records. This suggestion does not adequately account for how invasive that requirement would be. The observer would have to study the officer's phone screen to make sure that she is recording rather than texting her supervisor (or her spouse), for example. Alternatively, the officer would have to announce when she started and finished her recordkeeping to the observer (likely a university student), as though to a supervisor. Without this requirement, the observer will already have a difficult task in participating in the ride-along, documenting his experience with the officers, and making the officers feel comfortable enough that the documented experience will be candid and representative. That task (an important source of the study's information) would be made much harder by a requirement that the observer record the time and effort of recordkeeping.

5. The Proposal's Provision that the Monitor Team Assess the Legality of Encounters Is Appropriate.

The Proposal provides that the legality of observed encounters will be assessed by the Monitor's team. (ECF No. 660-1 at 6.) Plaintiffs argue that the group making these assessments should be "independent of the Monitor team, the NYPD and the Plaintiffs" and should "include as many members with civil rights and/or criminal defense backgrounds, including individuals who come from or work directly with communities that have been directly impacted by the NYPD's unconstitutional stop-and-frisk and trespass enforcement practices, as there are members with law enforcement and/or prosecutorial backgrounds". (ECF No. 669 at 3.) They assert that the Monitor's team is not diverse enough to make assessments that will be accepted as legitimate "in the eyes of community stakeholders, such as [CPR]" because "all but one" of the team members "are current or former police officers and/or prosecutors lacking a criminal defense and/or civil rights background". (*Id.*) Plaintiffs' arguments ignore the independence and diversity of the Monitor team, and their proposed modification to the Proposal would enhance neither the accuracy of the study's legality assessments nor the study's legitimacy. These arguments are further addressed by the fact that the Monitor recently added three former judges to his team who have expertise in the law related to stop, question and frisk and will help with the work of the pilot program.

A. Plaintiffs' Argument that the Monitor Team Cannot Impartially Assess the Observed Encounters Relies on Improper Premises.

Plaintiffs rely on the premise that they, the NYPD, and the Monitor team are all interested stakeholders in this process and that a new group of people must be recruited to get an unbiased perspective on the study's observed encounters. Yet the

parties are interested stakeholders; the Monitor team is not. And that is why this Court has trusted the Monitor with a special advisory role throughout this process, including in advising the Court whether the benefits of the proposed reforms under discussion ultimately outweigh the costs. (ECF No. 619 at 3.) Plaintiffs' and CPR's disagreements with the Monitor about a few points in the Proposal do not in any way suggest that the Monitor team is not independent, or that its Proposal would produce illegitimate results.

Plaintiffs also rely on the improper premise that the members of the Monitor team bring a pro-government bias to the observed encounters because they have worked as police officers or prosecutors. As an initial matter, Plaintiffs underestimate the diversity of experience that the Monitor team brings to its work, including significant civil-rights and criminal-defense experience. A chart summarizing some of that experience is attached as Exhibit 3.[11] More importantly, Plaintiffs' argument makes the surprising assumptions that everyone who works a particular kind of job has a similar perspective about the subject of the job, and that a single job in a person's background determines her perspective for life. Not all current or former prosecutors take a uniform, pro-government position in matters of criminal justice, and many one-time prosecutors go on to serve as judges, where their impartiality in all cases, including criminal cases, is of utmost importance. Plaintiffs' narrow view of the Monitor team should be rejected.

Professors Fagan and Glaser argue that we should think about the group assessing the legality of the observed encounters the way we think about an appellate

[11] Further detail about the experiences of the members of the Monitor team can be seen in the résumés attached as Exhibits 4 through 14.

panel, (ECF No. 669-1 at 2-3), and that analogy helpfully explains why the Proposal's existing setup is sound. An appellate panel may consist of judges who are all nominees of one party, or judges who are all former prosecutors—yet the makeup of the panel is not altered when it turns out to be composed that way. Instead, judges are trusted to administer the law impartially, taking seriously the duty of fairness with which the law charges them. This Court has charged the Monitor team with impartiality throughout this process and the team has faithfully executed that obligation. It should be trusted to fairly assess the legality of the observed encounters.

Plaintiffs' suggestion that the assessments be made by a panel comprising an equal number of members with backgrounds in law enforcement and criminal defense would not enhance the accuracy of those assessments. Furthermore, the Monitor team's assessments will not create any appearance of partiality, as the Monitor was appointed by this Court to act impartially and has done so throughout this assignment. Finally, the Monitor team's impartiality has been reinforced by the Monitor's recent decision to manage the workload of the pilot study by adding three former judges to the team. From their long experience formally assessing police-citizen encounters, these former judges bring not only expertise but a proven tradition of fairness to the study.

B. The Monitor Team Is Best Equipped to Assess the Legality of Police-Citizen Encounters Because of Its Significant Experience Making Such Assessments.

The flaws in Plaintiffs' premises that the Monitor team is not impartial and that particular jobs in the backgrounds of the team members determine their perspectives should lead this Court to reject Plaintiffs' argument. Moreover, the tasks that the Monitor

team already performs provide additional support for finding that the Monitor team is best equipped to make the study's legality assessments.

For example, the Monitor team examines the legality of stops assessed by the NYPD's Quality Assurance Division ("QAD"). The QAD reviews a sample of stop reports to see whether they articulate reasonable suspicion for the corresponding stops, reasonable suspicion for any ensuing frisks, and adequate justification for any ensuing searches. The QAD creates a record of its findings for each report. Every quarter of each year, the Monitor team then reviews a sample of the same set of stop reports that QAD examined. This sample is usually about 15% of the sample reviewed by QAD, or around 300 stop reports. Three members of the team assess the legality of the activities in each stop report; all disagreements are subject to group discussion and, if unresolved, are elevated to the Monitor and Deputy Monitor for review. The team's determinations on each report are completely independent of the same determinations made by the QAD; only after the team has completed its review of the reports does it check whether QAD reached the same conclusions. The team then meets with the QAD to discuss the reasons for their differences in order to motivate future improvement. The team has consistently found more activities illegal in the stop reports it reviews than the QAD did in reviewing the same reports.

Having the Monitor team review and assess the observed encounters puts the Monitor team in the same position it will be exercising in the BWC pilot program required by this Court's Remedial Order. Under that order, the Monitor is required to establish procedures for, among other things, assessing the effectiveness of BWCs in reducing unconstitutional stops and frisks and, at the end of the one-year pilot,

determining whether "the benefits of the cameras outweigh their financial, administrative, and other costs." *Floyd v. City of New York*, 959 F. Supp. 2d 668, 685 (S.D.N.Y. 2013). This Court did not require that the BWC pilot be overseen or that its results be evaluated by an independent panel of experts, nor have Plaintiffs expressed any reservations or objections to the Monitor team fulfilling its role in that pilot.

In that role, the Monitor team will be assessing the impact of cameras on police lawfulness (in addition to reviewing data that relate to the impact of cameras on the civility of police-citizen encounters and on the *DeBour* level of police activity). Stop reports provide an opportunity to examine whether cameras affect the lawfulness of police interactions with citizens. Each quarter, since the beginning of the BWC pilot, the Monitor team has been reviewing stop reports to assess whether NYPD officers comply with the Constitution and the provisions of the *Floyd*, *Ligon*, and *Davis* orders, and to compare whether the level of compliance differs between camera and non-camera precincts. The team will ultimately review a sample of stop reports large enough that it will be representative of stop reports made in the 20 pairs of precincts in the experiment. The team will then determine whether the presence of cameras influenced the officers' justifications for the stops, frisks, and searches, and also whether wearing cameras affected the demographic makeup of those stopped. There is no reason why the work of assessing encounters for their legality should be done by two different teams under the BWC pilot and the Proposal.

In addition to reviewing stop reports as part of its assessment of the work of the QAD and for the BWC pilot, the Monitor team determines the constitutionality of police-citizen encounters in other circumstances, too. The Monitor team has made those

determinations many times when reviewing the Citizen Complaint Review Board's SQF investigations and the NYPD's handling of the related cases. The Monitor team has also reviewed BWC video footage associated with stops. Going forward, the Monitor team will be reviewing and evaluating significant numbers of BWC videos as part of the monitoring required by the *Davis* and *Ligon* cases. The Monitor team will also be reviewing trespass arrests (and their associated records and videos) for their constitutionality and will be reporting on those reviews in future reports.

6. The Proposal Appropriately Accounts for the Views of the Impacted Communities.

The Monitor team is charged with developing the protocols that will govern the activities of observers and those assessing the legality of encounters. The parties and named stakeholders, including CPR, will have roles in this development consistent with the roles they have had in the development of the remedial process's past reforms.

Plaintiffs argue that the Proposal should include "a formal structured role for multiple representatives of impacted communities and/or members of organizations that work directly with those communities, such as CPR," in the design of these protocols. (ECF No. 669 at 4.) They especially want CPR's feedback on the "thorny legal questions" surrounding police-citizen encounters and implicated by the protocols. (*Id.* at 3.) This argument overlooks both the role that CPR has played in the design of the study and the fact that most of the legal judgments involved in these protocols have been settled and approved by this Court in the NYPD's SQF policies and training. Thus, the Proposal should not be changed on account of this argument.

Dating back to the Remedial Order, this Court has consistently provided that the reforms in this process are to be developed by the Monitor working in conjunction with the parties. (ECF No. 372 at 12.) This Court's order for the development of this Proposal followed that longstanding practice. (ECF No. 619 at 2.) This longstanding practice makes sense as a way of keeping the reform process manageable and ensuring that the timing of the process stays on track. Though it is not a party, CPR has been allowed to participate in this process by reviewing and commenting on the materials received by Plaintiffs. Although Plaintiffs have always had to request appropriate consent to share confidential materials with CPR, that consent has never been denied. (*See* ECF No. 647 at 5.)

When actual drafts of the protocols for legal assessors and observers are circulated to Plaintiffs, CPR will have the opportunity to review and comment, to the extent that Plaintiffs request and receive consent to show those drafts to CPR. Plaintiffs have not adequately explained why CPR requires a role in the development of these protocols more significant than that it has had in every other part of the remedial process.

Furthermore, the issue for which Plaintiffs chiefly request CPR's input—the "thorny legal questions" surrounding police-citizen encounters—has mostly been resolved by the SQF policies and training program, with the customary amount of participation from CPR. The SQF policies and training program were developed pursuant to the usual procedures: the Monitor and parties developed them, and CPR had the opportunity to comment on any materials that Plaintiffs decided to show to it. This Court approved the SQF policies on March 25, 2016. (ECF No. 247.) The training program covers "the fundamental principles of stop, question and frisk, trespass

enforcement, and bias-free policing". (ECF No. 615 at 1.) This Court approved the program on July 10, 2018. (*Id.* at 4.) Since that time, the program has been used to educate thousands of NYPD officers on the appropriate way to initiate and conduct encounters with citizens. When the Monitor team assesses the legality of officers' encounters in this study, it will assess whether the officers conducted themselves the way they were trained. In other words, the protocol for assessing the legality of encounters will be derived from the SQF policies and training program; no newly developed criteria are necessary or appropriate.

CPR goes further than Plaintiffs and asks for a role in the study greater than the parties' role. Among other things, it wants to "be provided with all of the data generated from the pilot program" (which will not be available to the NYPD or Plaintiffs) and "participate with the Monitor's team of experts" in legal assessments (which the NYPD and Plaintiffs will not do). (ECF No. 673 at 2-3.)

The basic problem with CPR's request is the same as that with Plaintiffs'. Neither has shown that this Court's longstanding practice concerning CPR's participation in this remedial process has eroded the legitimacy or reliability of the reforms the process has achieved. And neither has adequately explained why a special exception to that practice is warranted in this study. The reliability and legitimacy of the study is secured because it will be led by the independent, expert Monitor team.

7. The Proposal Already Assesses the Effect of Observer Presence on Officers.

The Proposal collects data from officers subject to each of the four test conditions who are not under observation. All of the officers in precincts and PSAs that are subject to condition 2, including unobserved officers, will be required to activate their

BWCs in Level 1 encounters. (ECF No. 660-1 at 2-3.) All of the officers in precincts and PSAs that are subject to condition 3, including unobserved officers, will be required to record Level 1 and 2 encounters on their phones. (*Id.*) And all of the officers in precincts and PSAs that are subject to condition 4, including unobserved officers, will be required to do both things. (*Id.*)

Furthermore, in order to test whether the presence of observers affects officer behavior, the study will compare data on the activity of officers in the presence and out of the presence of observers. (*Id.* at 14, 9.)

Therefore, CPR's argument that the Proposal should account for the effects of observer presence on officer behavior is moot. (ECF No. 673 at 4.)

8. The Proposal's Provision for the Publication of Data Is Appropriate.

The Proposal contemplates publishing aggregate data from the study in an appendix to the Monitor's final report. CPR argues that instead of waiting for this publication, the NYPD should be required to publish quarterly and annual reports on Level 1 and Level 2 encounters containing ten different types of information for each encounter. (ECF No. 673 at 5.)

The existing plan to wait for publication until study's end, at which point all data are in and the experts may examine and situate those data together, is the better one. (Indeed, it is the same approach that is being used for the current BWC pilot.) The Monitor will provide the parties with information on the status of the pilot as it proceeds, but requiring reports on an incomplete data set is more likely to sow confusion than improve understanding. More importantly, premature publication of the data could

contaminate the study, affecting officers' perceptions of the study's ongoing findings and so influencing officer conduct. This is why the best practice for scientific studies like this one is to wait until completion before publishing data.

* * *

For the foregoing reasons, the Monitor respectfully requests that this Court adopt the version of the Proposal attached as Exhibit 1 to this letter, which reports results in odds ratios and clarifies that officers will be required to document encounters in a timely manner, as discussed above.

Benjamin Gruenstein
CRAVATH, SWAINE & MOORE LLP
Worldwide Plaza
825 Eighth Avenue
New York, New York 10019
Phone: (212) 474-1000
Fax: (212) 474 3700
bgruenstein@cravath.com

Hon. Analisa Torres
United States District Court
Southern District of New York
500 Pearl Street
New York, New York 10007-1312

BY ECF